2Ш

УДК 519.7:007.52

О.Ю. Шевченко, М.В. Климова
Харківський національний університет радіоелектроніки, м. Харків, Україна
shevchenko@vk.kh.ua, klymova@gmail.com

Метод побудови інтелектуальних систем обробки інформації та документообігу за допомогою онтологічної бази знань

Стрімкі темпи науково-технічного прогресу потребують новітніх технологій збереження та обробки інформації для менеджменту різноманітними ресурсами, в тому числі в галузі освіти та науки. Для вирішення цієї задачі нами було застосовано онтологічний підхід і створено «Онтологічний портал менеджменту та оцінки освітніх і наукових ресурсів України». Цей портал покликаний виявляти приховані знання предметної галузі, встановлювати протиріччя, контролювати відповідність дій навчальних закладів до законодавства України та вимог суспільства.

Вступ

Стрімкі темпи науково-технічного прогресу потребують новітніх технологій збереження та обробки інформації для вирішення задач менеджменту різноманітними ресурсами, в тому числі в галузі освіти та науки України. Класичні системи менеджменту мають ряд недоліків, пов'язаних з великими складностями при зміні чи розширенні структури даних цих баз за необхідності пошуку різноманітної інформації або при інтеграції різних інформаційних баз.

Створення системи, заснованої на онтологічному підході, вирішує низку зазначених проблем. Головною перевагою онтологічних систем є гнучкість, здатність інтегрувати різноманітну семантичну інформацію у загальній базі знань, здатність автоматично структурувати цю інформацію, що надає можливості для інтелектуальної поведінки програмної системи, а саме: автоматичної інтеграції нових знань у загальну семантичну інформаційну структуру, виконання семантичного пошуку знань, виконання автоматичного тлумачення накопичених знань, перевірки на несуперечливість та достовірність інформації з онтологічної бази, виходячи з явних та неявних знань, які можуть накопичуватись в процесі функціонування систем. Важливою особливістю онтологічної бази знань є можливість автоматичного навчання системи та ефективного накопичення знань – процесу, при якому до бази знань системи зберігаються знання, придатні для розв'язання окремих завдань.

Мета даного дослідження – організація ефективного та прозорого менеджменту знань, що є гетерогенними та розподіленими за своєю природою, заснованого на онтологічному підході.

Постановка задачі дослідження

Задачею дослідження є розробка методу створення інтелектуальної системи для збереження інформації про організаційні, людські, інформаційні, матеріально-технічні ресурси України під управлінням Міністерства освіти, а саме семантичного загальнонаціонального порталу, який здатен зберігати великі обсяги інформації, зокрема

про освітні і наукові ресурси України, розробити механізми безперервного накопичення знань, видобування нових знань, тобто процесу виявлення джерел і самих знань з існуючої інформації, структурування знань, що полягає у виявленні основних понять галузі та розробки структури подання інформації, що їх стосується, обробки інформації, яка зберігається на порталі, забезпечивши підтримку виконання таких загальнонаціональних процедур контролю якості освіти, як акредитація та ліцензування, виконати формалізацію та програмну реалізацію системи та закласти механізми корекції існуючої інформації.

Сьогодні дані, що представляють та характеризують освітні та наукові установи та підрозділи, зберігаються в різних місцях, у відмінному один від одного вигляді та форматі, що призводить до відсутності можливості ефективної роботи з ними, саме тому першочерговою задачею є створення модуля накопичення інформації про ресурси Міністерства освіти і науки України шляхом їх реєстрації на порталі. Важливою вимогою до системи є можливість зручного керування та навігації накопиченими ресурсами, а також прозорість та відкритість системи. Структура сховища, що зберігатиме інформацію про ресурси, має бути зручною для обробки та гнучкою, оскільки однією з важливих умов є можливість швидкої адаптації до змін у структурі освітньо-наукової галузі та універсальність підходу з огляду на необхідність роботи з різними типами ресурсів.

Організація взаємодії компонентів онтологічної системи менеджменту

Для розробки інтелектуальної системи нами було обрано онтологічний підхід, що передбачає наявність онтології як бази знань, що зберігає як знання про ресурси Міністерства освіти і науки, так і знання про портал – його структуру, інтерфейс тощо.

Онтологічну систему можна охарактеризувати як комплекс засобів для накопичення знань, системи виконання запитів до вибірки, модифікації і поповнення знань та системи перевірки на несуперечливість та достовірність цих знань. Основну схему онтологічної системи наведено на рис. 1. Під комплексом засобів для накопичення знань ми розуміємо онтологічне сховище семантичних знань, побудоване на засадах таких технологій, як Sesame RDF [1], Protege [2] та інших, здатних накопичувати знання, подані у рамках стандарту RDF та OWL. Система виконання запитів для вибірки, модифікації і поповнення знань – це додаткові бібліотеки, які здатні виконувати запити до цих сховищ за допомогою SQL-подібних виразів, наприклад, мовою SparQL [3], чи за допомогою більш гнучкого та інтелектуального стандарту, такого як SWRL [4]. Як правило, реалізація цих систем базується на використанні бібліотеки Jena [5]. Системи перевірки на несуперечливість та достовірність знань, робота над якими активно ведеться провідними науковими лабораторіями [6], у нашій роботі реалізовано як систему контролю цілісності та достовірності онтологічної бази. Систему менеджменту та оцінки освітніх і наукових ресурсів України реалізовано у вигляді семантичного порталу, який містить явне (у вигляді онтології) подання власної структури й моделі предметної галузі та реалізує збір і публікацію інформації про галузь у форматі Semantic Web.

Система складається з набору проектів, призначених для розв’язання різних задач:

– системи менеджменту ресурсів – для опису та обчислення національних освітніх ресурсів в реальному часі;

– системи архівації інформації, що дозволяє отримати інформацію про національні освітні ресурси за підсумками попередніх років;
– системи підтримки процедур акредитації, що дозволяє підготувати потрібну документацію для акредитаційної справи.

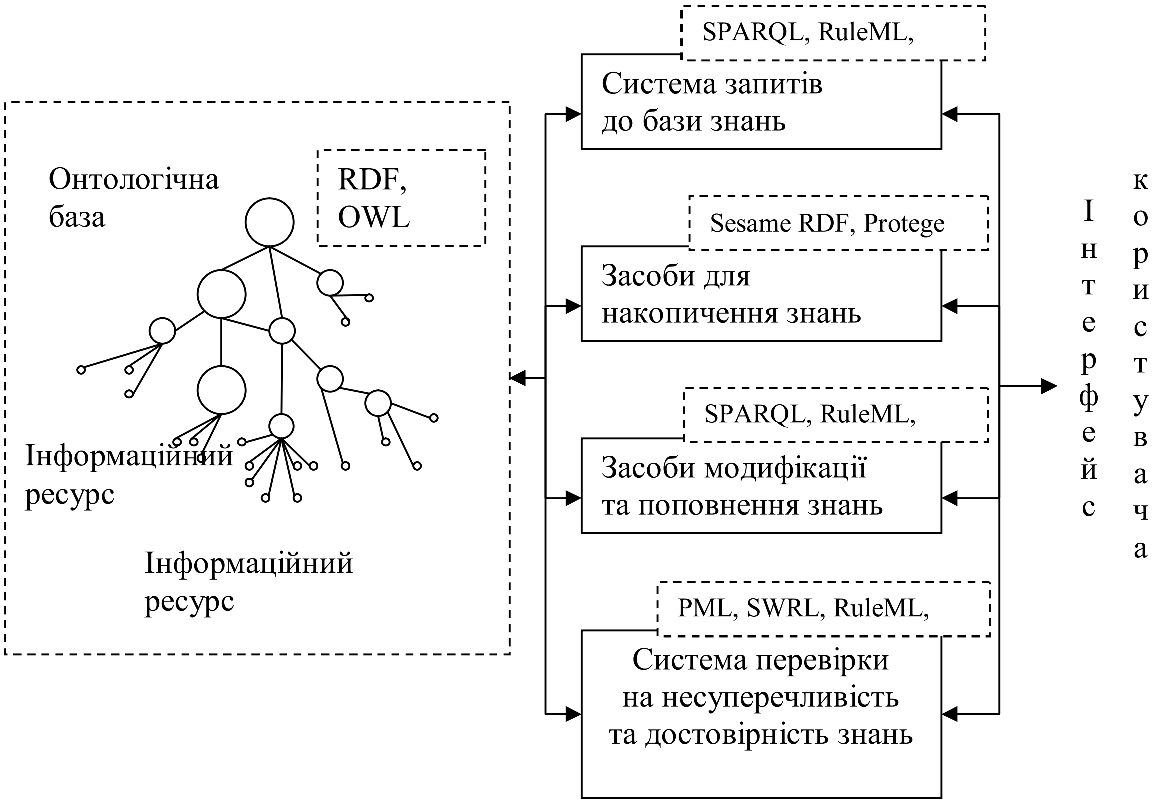


Рисунок 1 – Загальна схема побудови онтологічної інформаційної системи

Онтологія системи менеджменту та оцінки національних ресурсів України в галузі освіти та науки

Основою розробленої системи є вбудована в неї онтологія, яка описує концептуалізацію предметної галузі, зберігаючи формалізовані знання про освітньо-наукові ресурси, та дозволяє розв’язувати задачі, які поставлено перед системою. Ресурсами називаються всі об’єкти галузі, які задіяні в освітньонауковому процесі, впливають на його якість та підлягають реєстрації з метою подальшого аналізу та оцінювання. Такими об’єктами можуть бути як заклади різного типу – навчальні, наукові, так і люди, які в них працюють, студенти, які в них навчаються, курси, які викладаються, та інше. На рис. 2 відображено фрагмент онтологічної бази знань системи «Онтологічний портал менеджменту та оцінки освітніх і наукових ресурсів України», що містить формалізовані знання про ресурси різних рівнів освітньо-наукової ієрархії, об’єкти якої об’єднані відношенням агрегації, найвищими серед них є об’єкти класу Міністерство, а найнижчими – об’єкти класу Навчальна група. Всі описані в онтології ресурси прив’язані до одного з рівнів цієї ієрархії.

Організація внутрішньої структури порталу та його інтерфейсу заснована на розробленій системі інтеграції елементарних одиниць, описаних в онтології, – полів. Поле вказує на певний показник, що характеризує той чи інший ресурс в галузі осві-

ти і науки та який в тому чи іншому вигляді буде доступний користувачам порталу. Поле характеризується приналежністю до певної сторінки, тобто переліком ідентифікаторів, що визначають сторінки, на яких відбувається відображення конкретного поля, переліком користувачів, що мають право на його перегляд та редагування, та іншим. Розрізняють поля, значення яких обчислюється автоматично, та поля, значення яких вносяться користувачами вручну до онтології системи через web-інтерфейс порталу та можуть бути літералами простих типів даних, – прості поля, та поля, значення яких можуть бути об’єктами інших класів онтології (як правило, класу Ресурс), – комплексні поля. Прикладом комплексного поля є поле Викладач кафедри штучного інтелекту, значенням якого буде певний об’єкт класу Викладач, а прикладом простого поля є рік народження викладача. Значення полів, що обчислюються автоматично, не може редагуватись, а вираховується чи виводиться в результаті застосування певної семантичної або математичної формули, яка зберігається в онтології. Формула подається як множина елементів-змінних, що пов’язані між собою певним відношенням, яким ставляться у відповідність конкретні об’єкти класу Ресурс. За допомогою внутрішнього механізму інтерпретації формули відбувається розрахунок значення поля. Інформація, що вноситься користувачем до онтології шляхом реєстрації необхідних даних через web-інтерфейс системи або виводиться в результаті застосування формули, відображає систему параметрів, які характеризують ресурси МОН України та є кількісними чи якісними за своєю формою.

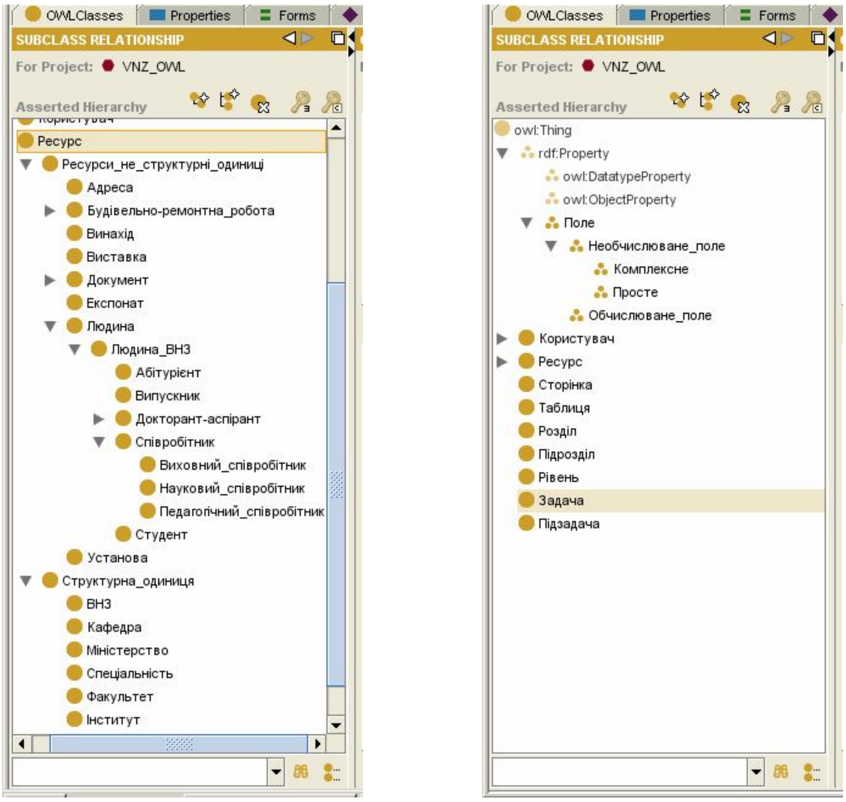


Рисунок 2 – Фрагмент внутрішньої онтології системи

Внутрішні елементи системи виконують автоматичне перетворення введеної інформації згідно зі стандартом OWL до семантичної анотації ресурсу, яка зберігається у спеціальному онтологічному сховищі знань та використовується під час виконання задач пошуку чи вибору інформації.

Таким чином, портал дозволяє в явному вигляді працювати з об’єктами предметної галузі. Структура порталу дозволяє, однозначно визначивши певний ресурс, відтворити всі зв’язки для нього та віднайти та виявити всі, навіть приховані, знання про ресурс, що існують в базі порталу. Якщо викладач працює на декількох кафедрах і ця інформація не є явно вираженою в його анотації, оскільки в анотації цього викладача було записано тільки одну кафедру або один університет, система дозволить віднайти неявну інформацію про всі місця роботи цього викладача. Це дозволяє вирішити проблему неоднозначності, нескоординованості та надлишковості знань, що існують в галузі освіти та науки та ускладнюють процеси, що відбуваються в цій галузі, а надто процеси контролю за якістю.

Динамічне формування структури та інтерфейсу системи

Використання онтологічної бази знань надає гнучкості побудованій на ній програмній системі. Онтологія здатна зберігати не тільки знання про предметну галузь, але й про структуру системи. Ця властивість застосовується для створення поліморфних систем, здатних змінювати свій зовнішній вид, що дозволяє добитися вищого рівня абстракції у розробці програми та адаптувати систему під кожного з її користувачів. Особливістю створеної онтології порталу є те, що вона зберігає знання не тільки про ресурси галузі, але й про структуру та інтерфейс порталу через множину так званих «інтерфейсних» класів, що відповідають за опис побудови структури порталу, як наприклад, класи Сторінка, Таблиця, Поле таблиці та інші (рис. 3), схема взаємодії між якими вказує на структуру інтерфейсу.

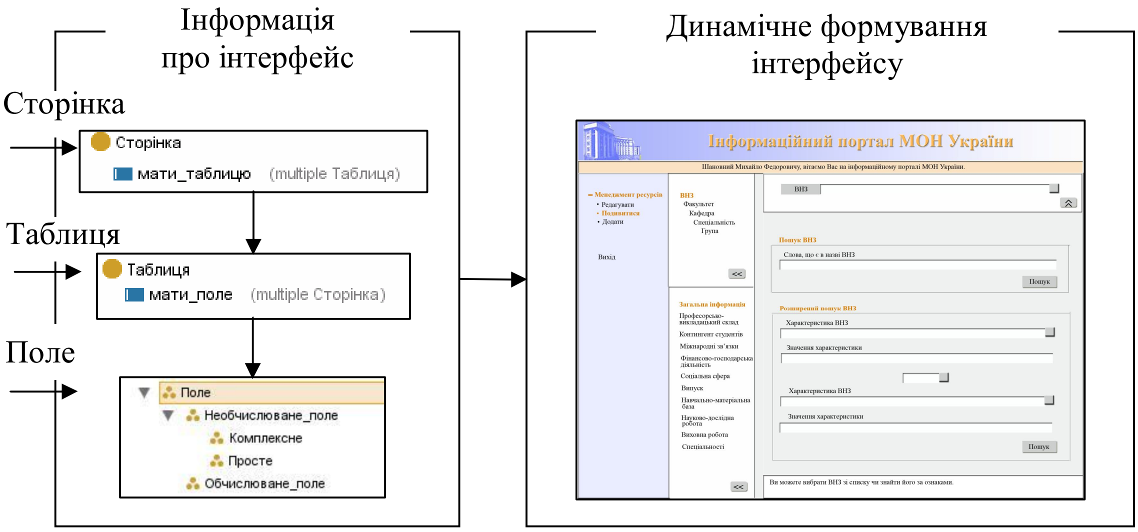


Рисунок 3 – Динамічне формування інтерфейсу

Для кожного з класів, що формують інтерфейс, закладено зв’язок з об’єктами типу Ресурс, що забезпечує автоматичне заповнення структури порталу даними про ресурси Міністерства освіти і науки. Така організація порталу дозволяє говорити про динамічне формування інтерфейсу. Інтерфейс онтологічного порталу складається з множини блоків, між якими відбувається взаємодія та обмін даними, – Блок задач, Блок рівнів, Блок розділів, Блок ідентифікації, Блок контенту та Блок підрозділів. Кожен блок отримує своє наповнення з онтології та здатен обмінюватись ним з іншими блоками.

Для кожного користувача зміст кожного з блоків є унікальним. Інтерфейс порталу та заповнення його конкретними даними відбувається згідно з правами користувача, розподіл яких реалізовано за групами або категоріями. Приналежність користувача до певної категорії визначається в першу чергу його посадою. Кожен користувач має доступ до перегляду або редагування ресурсів тільки свого рівня, який відповідає тій структурній одиниці, до якої належить користувач. Рівні, до яких має доступ користувач, відображено в Блоці рівнів та представлено класами в онтології. Для кожного ресурсу в онтології в явному вигляді описані категорії користувачів, що мають право на його перегляд та редагування.

Користувач може мати дозвіл на те, щоб бачити ту інформацію, що стосується його рівня, але яка була сформована на нижчому рівні, але доступ до редагування цієї інформації він не матиме. Так, наприклад, секретар кафедри матиме доступ тільки до інформації, яка стосується кафедри, до якої він належить. Головною проблемою розподілу прав користувачів лишається керування правами тих користувачів, які належать до декількох категорій різних рівнів. Механізми, покладені в основу розподілу прав користувачів та динамічного формування інтерфейсу та змісту порталу, гарантують безпеку, гнучкість та об’єктивність роботи порталу.

Забезпечення несуперечливості та достовірності знань

У випадку поєднання мови запитів та середовища, що надає відповіді на запитання як результат складних та комбінованих кроків дедуктивного виведення або обчислень, що є характерним для інформаційних систем, необхідною є розробка модуля перевірки на несуперечливість та достовірність отриманих знань, який формував би пояснення та підтвердження процесів отримання знань. Структура порталу забезпечує побудову достатньо відкритого та прозорого середовища для менеджменту знань.

Організація онтології дозволяє отримати контроль за наявністю та якістю будь-якого ресурсу галузі, оскільки для обчислення кожного з ресурсів в системі необхідне посилання на цей ресурс або на електронний документ, що підтверджує його наявність та відповідність, згідно з процедурами верифікації, які забезпечують цілісність та достовірність інформації. Застосування онтології надає можливості автоматичної перевірки отриманих результатів на суперечливість або нецілісність шляхом перевірки транзитивних та рефлексивних властивостей класів онтології.

Процедури обчислення показників спираються на інформацію, представлену в семантичних анотаціях ресурсів та сформовану автоматично при обчисленні показників вищого рівня. Доказ відповідності показників містить інформацію про джерело отриманих знань у вигляді семантичної анотації, що містить метадані джерела, та інформацію-пояснення про процес виведення, який використовувався для генерації отриманого результату. Кінцевий доказ, який надається користувачу порталу, можна представити у вигляді дерева, гілками якого є кроки виведення певного результату, що пояснює процес дедукції наслідку з певного набору передумов. Дерево виведення результату складається з фрагментів доказів, об’єднуючи які, можна простежити повний шлях виведення певного наслідку від початкового стану середовища – сукупності кроків виведення, у яких передумови були задані конкретним користувачем. Основним типом правил, за якими відбувається виведення значень показників в порталі, що характеризують ресурси, є формула, елементам якої у відповідність ставляться певні об’єкти предметної галузі. Таким чином, створено повністю прозору систему формування параметрів освітнього процесу.

Висновки

У статті запропоновані: метод побудови інтелектуальних систем обробки інформації та документообігу за допомогою онтологічної бази знань, метод динамічного формування структури та інтерфейсу системи та метод забезпечення несуперечливості та достовірності знань, також запропоновано метод побудови систем, що надає необхідну гнучкість, універсальність, захищеність інформації та можливості для легкого її розширювання. На прикладі показано, що застосування онтологій як бази знань для зберігання інформації про освітньо-наукові процеси дозволяє досягти прозорості у формуванні та обробці ресурсів цієї галузі, а застосування web-орієнтованого інтерфейсу дозволяє зробити систему розподіленою, відкритою та легко доступною.

Література

1. Jeen Broekstra. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema / Jeen Broekstra, Arjohn Kampman, Frank van Harmelen // International Semantic Web Conference 2002. – Sardinia (Italy) [Електронний ресурс]. – Режим доступу : http://www.openrdf.org/doc/papers/Sesame-ISWC2002.pdfs
2. Creating Semantic Web Contents with Protege-2000 / N.F. Noy, M. Sintek, S. Decker, M. Crubezy, R. Fergerson, M.A. Musen // IEEE Intelligent Systems. – 2001. – P. 16, 2, 60-71.
3. SPARQL Query Language for RDF. W3C Recommendation [Електронний ресурс], 2008. – Режим доступу : http://www.w3.org/TR/rdf-sparql-query/
4. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission / Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean [Електронний ресурс], 2004. – Режим доступу : http://www.w3.org/Submission/SWRL/
5. Jena – A Semantic Web Framework for Java [Електронний ресурс]. – Режим доступу : http://jena.sourceforge.net/
6. William E. Perry. Quality Assurance for Information Systems: Methods Tools, and Techniques / William E. Perry // QED Information Sciences (Inc. Wellesley, MA, USA), 1991. – 814 pp.

А.Ю. Шевченко, М.В. Климова

Метод построения интеллектуальных систем обработки информации документооборота при помощи онтологической базы знаний

Стремительные темпы научно-технического прогресса требуют современных технологий сохранения и обработки информации для менеджмента разнообразными ресурсами, в том числе в области образования и науки. Для решения этой задачи нами использовался онтологический поход, был создан «Онтологический портал менеджмента и оценки образовательных и научных ресурсов Украины». Этот портал призван выявлять скрытые знания предметной области, устанавливать противоречия, контролировать соответствие действий учебных заведений законодательству Украины и требованиям общества.

Yu. Shevchenko, M.V. Klymova

The Building Approach of Intellectual Systems of Information Processing and Document Revenue by Means of Ontology Database

Swift rates of scientific-technical evolution demand modern technologies of information storing and processing for diverse resources management, for example in science and education domain. For solving of this task the ontological approach was used and “Ontology-based Portal for Management and Evaluation of scientific and Educational Resources of Ukraine” was created to reveal implicit domain knowledge, detect conflicts and control the correspondence of the HEI`s activity to the legislative base of Ukraine and society needs.

Стаття надійшла до редакції 12.05.2009.